

Education

Ph.D., Georgia Institute of Technology, School of Computer Science, Computer Science 01/22 - Present
B.A., University of Central Florida, Economics 12/11

Research Experience

SAIL: Systems for Artificial Intelligence Lab
Graduate Research Assistant 01/22 - Present

Professional Experience

Komodo Health - San Francisco, California
Senior Data Scientist, Machine Learning 10/20 - 07/21

- Developed Machine Learning (ML) systems using clinical data to detect rare and undiagnosed diseases
- Designed greedy optimization algorithms to solve class imbalance problems with sparse data to engineer near-optimal datasets for ML, resulting in 4x effectiveness over previous manual process
- Managed a ML engineering team focused on algorithm development, including mentoring of junior scientists
- Led complex cross-functional ML projects with pharmacological and clinical subject matter experts

Change Healthcare - Emeryville, CA

Senior Data Scientist, Artificial Intelligence 01/19 - 05/20

- Led end-to-end development of a Deep Learning (DL) based medical bill auditing system for scanning and analyzing >100M pages of paper clinical records, leading to significant reduction of waste, fraud, and abuse
- Built production and testable code for an enterprise Natural Language Processing (NLP) system capable of launching AWS pipelines to automate manual insurance claim auditing, generating \$400K in cost savings
- Managed team developing platform for information extraction and clinical predictions using a large corpora of unstructured text which enabled rapid and turn-key deployment of new models for our business partners
- Onboarded and managed a remote team of 5 medical annotation specialists, labeling 360K clinical records via programmed annotation tools to train DL models while saving \$200K versus alternative option
- Partnered with product management and medical scientists to define and prioritize team roadmap

LendUp - San Francisco, CA

Data Scientist 01/17 - 10/18

- Organized full development lifecycle of XGBoost and Logistic Regression models with Hyperopt for credit card underwriting, responsible for ideation, testing, validation, deployment, and iterative improvements
- Launched a new cross-functional team with Growth Marketing to develop ML platform using XGBoost
- Developed new customer micro-targeting when predicting marketing responses driving \$1.5M in revenue
- Eliminated redundant APIs calls made by underwriting ML model, saving \$110K annually
- Drove overhaul of Operations Analytics department's analytics capabilities, updating front end, rebuilding SQL, and updating reporting capabilities to revert years of SQL tech debt
- Rearchitected Operations Analytics Department including new user experience features, rebuilding data pipelines, and providing new automated reporting capabilities while eliminating years of technical debt
- Developed internal data science career program, providing technical mentorship for 20 employees across the company and successfully placing 4 employees in data analyst roles

Skills

Computer Science & Engineering Python, Shell Scripting, SQL/NoSQL, Ray, Modin, MLFlow, Jupyter Notebook, Airflow
Cloud Services (Amazon, Google, and Azure), Cluster Computing (Spark, Kubernetes)
Parallelized and Distributed Computing, Unit and Integration Testing, CI/CD

Machine Learning & Algorithms PyTorch, Scikit-learn, PySpark, Pandas, Dask, SciPy, NumPy, SpaCy, HuggingFace, XGBoost
Deep Learning, Neural Networks (Recurrent, Convolutional, Transformer, Graphs)
Machine Learning (Graphs, Decision Trees, Logistic Regression), Hierarchical Embeddings, Hierarchical Clustering, Natural Language Processing (NLP), Statistics and Probability

Honors & Awards

Research

UnfoldML: A 2-D Pipeline blah blah

Xu Y., Khare A., Matlin G., Monish R., Zhang C., Tumanov A., (2022)

Publications

Presentations

Grants

Certifications

Associations

References